Near-optimal Binary Compressed Sensing Matrix

Weizhi Lu, Weiyu Li, Kidiyo Kpalma and Joseph Ronsin

Abstract

Compressed sensing is a promising technique in signal processing that attempts to recover sparse signal through as few linear and nonadaptive measurements as possible. In this sense, the recovery performance is largely determined by the structure of measurement matrix. Recently, $\{0,1\}$ binary random matrix is found attractive for its low computation load as well as for its comparable performance with Gaussian Matrix. However, the optimal distribution of binary random matrix for compressed sensing remain unknown in theory. This paper, for the first time, deterministically defines one class of near-optimal binary random matrix by comparing the restricted isometry property (RIP) of regular binary random matrix characterized with the correlation between columns. In theory, the proposed matrix is expected to achieve nearly the best RIP with as few nonzero entries as possible. Simulation results also confirm the availability of the proposed near-optimal matrix, with better performance over both traditional binary random matrix and Gaussian matrix.

Index Terms

binary random matrix, compressed sensing, correlation, near-optimal, restricted isometry property, RIP.

I. Introduction

Suppose that a k-sparse signal $x \in \mathbb{R}^N$, with at most *k* nonzero entries, is measured by an undetermined matrix $A \subset \mathbb{R}^{M\times N}$ with $M < N$

$$y = Ax, \tag{1}$$

Manuscript submitted April 15, 2013.

Weizhi Lu, Kidiyo Kpalma and Joseph Ronsin are with Université Européenne de Bretagne (UEB), INSA, IETR, UMR 6164, F-35708 Rennes, France. Email: {weizhi.lu, kidiyo.kpalma, joseph.ronsin}@insa-rennes.fr

Weiyu Li is with CREST-ENSAI and IRMAR-Université Rennes 1, 35172 Bruz, France, and also with Institute of Mathematics, Shandong University, 25510, Jinan, China. Email:weiyu.li@ensai.fr

compressed sensing [1] states that one can recover x perfectly from finite observations $y \in \mathbb{R}^M$. This problem is customarily formulated as an l_1-minimization based convex optimization problem

$$\min ||\hat{x}||_1 \text{ subject to } y = A\hat{x} \tag{2}$$

provided that restricted isometry property (RIP) is satisfied. To understand RIP, we first review the concept k-restricted isometry constant (RIC) δ_k, which is the smallest quantity obeying

$$(1-\delta_k)||x_T||^2 \ll ||A_T x_T||^2 \ll (1+\delta_k)||x_T||^2 \tag{3}$$

for all subsets $T \subset \{1, ..., N\}$ with cardinality $|T| \leq k$ and all vectors $x_T \in R^{|T|}$, where A_T denotes the set of columns of A with indices in T. And then RIP of order k asserts that the k-sparse signal can be recovered faithfully by Eq. (1), if δ_{2k} is less than some threshold, i.e., 0.4652 in [2]. Roughly speaking, RIP requires that the symmetric matrix $A'_T A_T$ approximates isometry.

It is well known that some random matrices generated by certain probabilistic processes, like Gaussian or Bernoulli processes [1] [3], satisfy RIP with high probability. In terms of the works related to Johnson-Lindenstrauss lemma [4] [5], these dense matrices are allowed to be sparsely sampled without obvious performance loss in compressed sensing. However, due to the randomness in structure and the uncertainty on RIP, these random matrices are prohibited in real applications. Consequently, several deterministic sensing matrices based on some special codes, like Reed-Solomon codes [6] [7], Reed-Muller codes [8], chirp sensing codes [9] etc., are sequentially proposed. In spite of explicit structure, these matrices still suffer from uncertain RIP and unstable performance [10].

Recently, the sparse binary parity-check matrix defined by LDPC codes [11] (shortly called 'LDPC matrix' in this paper) attracts our attention for its competitive performance as well as circuit-friendly structure [12]. Up to now, the structure and construction for better LDPC matrix in coding theory have been widely studied. In contrast, the literature for better performance in compressed sensing is rare. In fact, as one class of $\{0, 1\}$ binary *sparse* matrix, LDPC matrix has no distinguishable characteristic in structure with its original definition [11], except for its particular representation and analysis methods based on bipartite graph. In this sense, the research for better LDPC matrix is essentially equivalent to searching more feasible sparse binary matrix with better performance while lower sparsity. However, to the best of our knowledge, there seems no impressive work to address this problem. In the earlier works [13] [14] [15] [16], LDPC matrix is only simply used as a common sparse binary random matrix [17], without particular evaluations on structure, i.e., *girth* or correlation between columns. Recently Dimkis et al. [18] propose that the *good* LDPC matrix in channel coding is probably the *good* sensing matrix in compressed sensing, by building a mathematical connection between channel coding and compressed sensing with linear programming decoding. Furthermore, they state that the *good* matrix should have *large* girth. Unfortunately, these theoretical results are still rough from the viewpoint of application. Most recently, Li et al. [10] evaluate one class of LDPC matrix based on Berlekamp-Justesen codes, and obtain deterministic RIP and comparable performance with Gaussian matrix. Obviously, all the aforementioned works are far away from the ideal goal of defining the better or even best sparse binary matrix in compressed sensing.

In this paper, by exploring the connection between the correlation and the *girth* of LDPC matrix, for the first time, we theoretically define one class of near-optimal $\{0, 1\}$ binary random matrix for compressed sensing, which achieves nearly the best RIP with nearly fewest nonzero entries. Significantly, in practice this kind of matrix can be approximately determined and constructed with progressive edge-growth (PEG) algorithm [19]. Numerical simulations also show that, the proposed near-optimal matrix indeed exists with better performance over both traditional binary random matrix and Gaussian matrix. In the following study, to distinguish with previous studies on binary matrix, we prefer to use 'LDPC matrix' to refer to the sparse binary random matrix characterized with the tool: bipartite graph, popularly used for LDPC codes.

Moreover, it is necessary to mention that the accurate calculation of RIP has been regarded as a difficult or even impossible assignment [20]. For instance, recent works [21] [22] have proved that the RIP solution to a given matrix with a given order k is NP-hard. However, the computational obstacle can be penetrated by analysis [23]. To search best binary random matrix, this paper will approximate the RIP of random matrix by analyzing its distribution. As will be detailed later, the solution to RIP can be derived by seeking the extreme singular values of A_T [23]. Unfortunately, the accurate or comparable extreme singular values of random matrix, cannot be well approximated with current random matrix theory [24]. Recall that the paper focuses on the comparison between RIPs rather than accurate RIP. Thus, in this paper the solution to the extreme singular values of A_T is roughly relaxed to calculating the extreme eigenvalues of symmetric matrix of similar distribution with $A'_T A_T$. In other words, $A'_T A_T$ is simply regarded as symmetric rather than positive semidefinitive. As it is known, the extreme eigenvalues of symmetric matrix has been widely studied in random matrix theory, and accurate solutions can be derived or approximated on some constraints.

The rest of the paper is organized as follows. In next section, the *regular* binary random matrix is defined by two types of LDPC matrices of different girth distributions. Relative concepts about Tanner graph and girth are also introduced. In section III, the RIPs of above two types of LDPC matrices are firstly calculated and analyzed in Theorems 1-3, and then the near-optimal binary random matrix are further derived by the comparison of RIPs in Theorem 4. In section IV, the empirical performance of proposed near-optimal matrix constructed with PEG algorithm is verified, by comparing with LDPC matrix with different sparsity as well as other popular matrices. Finally, we end this paper with a conclusion in section V.

II. Fundamentals of LDPC Matrix

In the study of LPDC codes, LDPC matrix tends to be represented by *Tanner Graph* (in *Definition 1*), and analyzed with *girth* (in *Definition 2*, abbreviated as g). According to the distribution of *girth*, this paper, for the first time, categorizes *regular* binary random matrix without same columns, into two types of LDPC matrices: LDPC matrix with girth $g > 4$ (in *Definition 3*) and LPDC matrix with girth $g = 4$ (in *Definition 4*). This novel definition will be helpful for the following computation of RIP. Currently, LDPC matrix with $g > 4$ can be constructed with numerous algorithms thanks to wide studies on LDPC codes. In contrast, it seems that there is no off-the-shelf knowledge to the construction and analysis of LDPC matrix with $g = 4$ due to its bad performance in LDPC codes. So this paper cannot sufficiently describe its structure in the following study. Typically, to express the variation of the correlation between columns with degree d varying, where d denotes the number of nonzero entries in each column, the columns of LDPC matrix are normalized, and thus nonzero entries are set to $1/\sqrt{d}$ instead of 1.

Definition 1 (Tanner Graph): A bipartite graph holds N variable nodes and M check nodes, respectively corresponding to N columns and M rows of binary matrix $\{0, 1\}^{M \times N}$. Variable nodes and check nodes are connected by the nonzero entries of binary matrix.

Definition 2 (Girth): In Tanner graph, the girth is defined as the length of the shortest cycle through all variable nodes.

Definition 3 (LDPC Matrix with girth $g > 4$): A binary matrix $A(M, N, d) \subset \{0, 1/\sqrt{d}\}^{M \times N}$, consists of $2 \leq d \leq M - 2$ nonzero entries per column and Nd/M nonzero entries per row. In structure, any two columns are allowed to share at most one same nonzero position. In adjacent Tanner graph, the girth is larger than 4.

Definition 4 (LDPC Matrix with girth $g = 4$): A binary matrix $A(M, N, d, s) \subset \{0, 1/\sqrt{d}\}^{M \times N}$, consists of $3 \leq d \leq M - 2$ nonzero entries per column and Nd/M nonzero entries per row. The largest correlation value between two distinct columns is s/d with $2 \leq s \leq d - 1$. In adjacent Tanner graph, the girth is equal to 4.

For the performance of sensing matrix, the correlation between distinct columns has been a fundamental index. Fortunately, the correlation of LDPC matrix with $g > 4$ can be statistically determined as shown in *Lemma 1*. Conversely, as the former states, it is hard to detail the correlation distribution of LDPC matrix with $g = 4$, due to limited knowledge. But one can roughly conjecture that its nonzero correlation values are taken from the set $\{1/d, ..., s/d\}$. Therefore, with the rough relation between the largest correlation and compressed sensing [25]

$$k < \frac{1}{2}(1 + 1/\mu) \tag{4}$$

where $\mu := \max_{i \neq j}\{|a_i' a_j|\} = 1/d$ or s/d for above two types of LDPC matrices, it is reasonable to expect that the larger d probably defines the better LDPC matrix. Also, this paper is developed with this inspiration.

Lemma 1 *(Correlation of LDPC matrix with $g > 4$)*: Any two distinct columns of LDPC matrix $A(M, N, d)$ in *Definition 3* take on correlation values

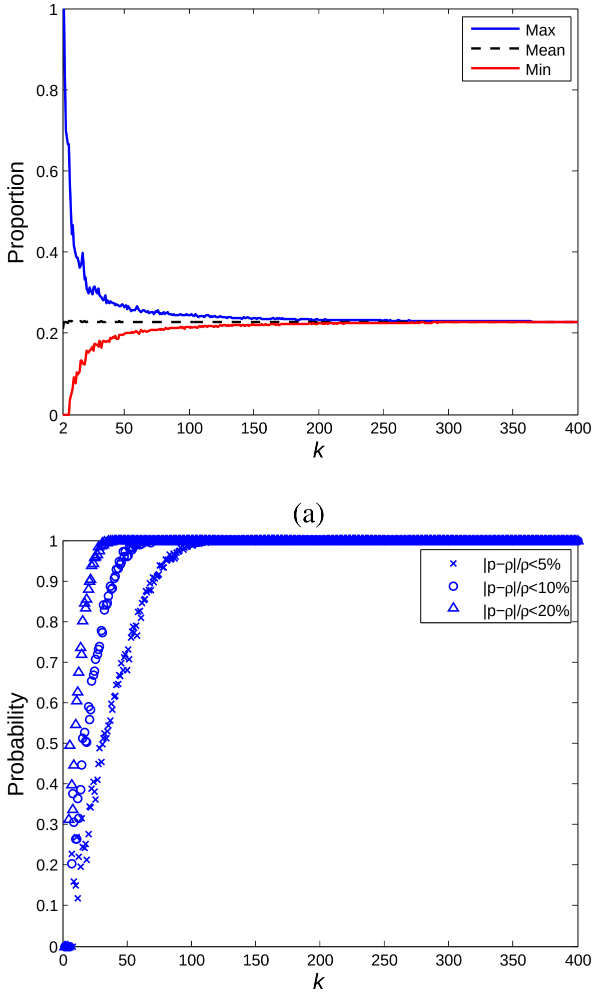
$$a_i' a_{j, j\neq i} = \begin{cases} 1/d & \text{with probability } \rho = \frac{Nd^2 - Md}{(N-1)M} \\ 0 & \text{with probability } 1 - \rho \end{cases} \tag{5}$$

where a_i and a_j denote two distinct columns of $A(M, N, d)$.

Proof: In Tanner graph associated with $A(M, N, d)$, any variable node v_i, $i \in \{1, ..., N\}$, holds d neighboring check nodes c_{b_k}, where the subscript $b_k \in C \subset \{1, ..., M\}$ denotes the index of check node, $k \in \{1, ..., d\}$, $|C| = d$; each check node c_{b_k} further connect with other $\frac{Nd}{M} - 1$ variable nodes v_j, where $j \in V_{b_k} \subset \{1, ..., N\} \setminus i$ represents the index of variable node, $|V_{b_k}| = \frac{Nd}{M} - 1$. Since variable node v_i has girth> 4 in *Definition 1*, we have that $V_{b_e} \bigcap V_{b_f} = \emptyset$, where $e, f \in \{1, ..., k\}$ and $e \neq f$, then derive $|V_{b_1} \bigcup V_{b_2} \bigcup ... \bigcup V_{b_k}| = d \times (\frac{Nd}{M} - 1)$. Therefore, among $N - 1$ variable nodes, there are $\frac{Nd^2 - Md}{M}$ variable nodes connecting with variable node v_i through one check node. Equivalently, any column of $A(M, N, d)$ has $\frac{Nd^2 - Md}{M}$ correlated columns with correlation value $1/d$. Then formula (5) is proved. ■

III. LDPC MATRIX FOR COMPRESSED SENSING

In this section, the RIPs for LDPC matrices with $g > 4$ and $g = 4$ are first derived and analyzed in Theorems 1-3, and then the near-optimal LDPC matrix is proposed and discussed by comparing RIP in Theorem 4.



(a)

(b)

Fig. 1. (a) Maximum, mean and minimum of the proportion of nonzero entries in the off-diagonal of $A'_T A_T$, $|T| = k$, for LDPC matrix $A(400, 200, 7)$. The mean is equal to the theoretical value $\rho = 0.2281$ in *Lemma 1*. (b) The probability that the proportion p of nonzero entries in the off-diagonal of $A'_T A_T$ centers on ρ with error bound $|p - \rho|/\rho$. Each point of (a) and (b) is derived with 10^3 simulations.

A. *RIP of LDPC matrix with girth larger than 4*

With the definition of LDPC matrix with $g > 4$, it is easy to derive that $A'_T A_T \subset \{0, 1, 1/d\}^{k\times k}$ with $T \in \{1, ..., N\}$ and $|T| = k$[1], is a positive semi-definite symmetric matrix with the diagonal equal to 1, and the off-diagonal of the distribution illustrated in *Lemma 1*. Theoretically, the solution to the

[1] Suffering from some possible confusion with the former notation in RIC, we simply assume $|T| = k$ in the following part.

RIC-δ_k of $A'_T A_T$, can be transformed to the pursuit for the extreme eigenvalues of $A'_T A_T$ (equivalently, the square of the extreme singular values of A_T), since

$$1 - \delta_k \leq \lambda_k \leq \frac{x'_T A'_T A_T x_T}{x'_T x_T} \leq \lambda_1 \leq 1 + \delta_k \tag{6}$$

where λ_1 and λ_k represent the two extreme eigenvalues of $A'_T A_T$. Note that, in this paper $\lambda_1 \geq \lambda_2 ... \geq \lambda_k$ are customarily used to denote the order of eigenvalues of $A'_T A_T$.

As the introduction states, it is hard to directly derive the extreme singular values of A_T, and so we turn to calculate the extreme eigenvalues of random symmetric matrix of similar distribution with $A'_T A_T$. In this paper, we prefer a simple algebra algorithm [26] rather than other popular results based on semicircle law [27], for the solution of extreme eigenvalues. The major advantage of the former is that it can provide an accurate solution for random symmetric matrix, if the random matrix could achieve some special distribution. In contrast, the accuracy of the latter can only be expected if the matrix size k is sufficiently large. Obviously, this is unfavorable for RIP usually with small k. In the following Theorem 1, the RIP-1 of LDPC matrix with $g > 4$ is derived with [26].

Theorem 1 (RIP-1): LDPC matrix $A(M, N, d)$ in *Definition 1* satisfies RIP with RIC

$$\delta_k = \frac{3k - 2}{4d + k - 2} \tag{7}$$

With $\delta_{2k} < 0.4652$ [2], further derive $k < 0.3671d + 0.2110$, for the faithful recovery based on l_1-minimization.

Proof: Please see Appendix A. ■

Note that, as the proof detailed in Appendix A, the two extreme eigenvalues for RIP-1 are achieved on the assumption that the proportion p of nonzero entries in the off-diagonal of $A'_T A_T$, could take value 1 or 0.5, for any $|T|$. However, as *Lemma 1* discloses, this condition cannot be satisfied all the time, because the proportion p should center on $\rho < 1$ with higher probability as $|T|$ increases. For better understanding, we further give an example in Figure 1, in which the proportion p for LDPC matrix $A(200, 400, 7)$ fast converges to the theoretical value $\rho = 0.2281 < 0.5$, as $|T|$ increases. In this case, for the large LDPC matrix with RIP of order k large enough such that $p = \rho$, its RIP-2 can be asymptotically evaluated with semicircle law [27], as shown in Theorem 2. Note that, Theorem 2 is only feasible for LDPC matrix with RIP of order k large enough such that the condition for semicircle law, $|T| \to \infty$, could be well approximated. Thus, considering generality and accuracy, RIP-1 rather than RIP-2 is used in the following comparison between RIPs.

Theorem 2 (RIP-2): Assume that the off-diagonal elements of $A'_T A_T$ take nonzero values with probability $\rho = \frac{Nd^2 - Md}{(N-1)M}$, for any $|T|(|T|-1)\rho \geq 2$, then drive that, the RIC-δ_k of $A(N, M, d)$ can be approximately formulated as

$$\delta_k = \frac{k\rho + 2\sqrt{k\rho(1-\rho)} + 1}{k\rho - 2\sqrt{k\rho(1-\rho)} + 3} \tag{8}$$

if $k = |T| \to \infty$.

Proof: Please see Appendix B. ■

B. *RIP of LDPC matrix with girth equal to 4*

For LDPC matrix $A(N, M, d, s)$ with $g = 4$ and the largest correlation $\mu = s/d$, the off-diagonal of $A'_T A_T$ possibly takes values from the set $\{0, 1/d, ..., s/d\}$, where $3 \leq d \leq M-2$ and $2 \leq s \leq d-1$. With the same solution algorithm [26] to Theorem 1, the RIP-3 of LDPC matrix with $g = 4$ is derived in Theorem 3.

Theorem 3 (RIP-3): LDPC matrix $A(M, N, d)$ with $g = 4$ and $\mu = s/d$, where $2 \leq s \leq d-1$ and $3 \leq d \leq M-2$ satisfies RIP with RIC

$$\delta_k = \begin{cases} \frac{(3k-2)s}{(k-2)s+4d} & if\ 3 \leq d \leq \frac{M}{2} \text{ and } 2 \leq s \leq d-1 \\ \frac{(3k-2)s+(k-2)(M-2d)}{(k-2)s-(M-2d)k+2M} & if\ \frac{M}{2} < d \leq M-2 \text{ and } 2d-M \leq s \leq d-1 \end{cases} \tag{9}$$

Proof: Please see Appendix C. ■

Note that, like RIP-1, the two extreme eigenvalues for RIP-3 are also achieved as the off-diagonal elements of $A'_T A_T$ can take the maximal nonzero value s/d with probability 1, or take binary value $\{0, s/d\}$ with equal probability. Nevertheless, as the section II states, in practice it remains unknown whether the above distributions could be well satisfied merely with limited knowledge about LDPC matrix with $g = 4$. And so this paper cannot ensures that the extreme eigenvalues for RIP-3 possibly are surely achieved by the practical matrix.

C. ***Our contributions***: *Near-optimal binary random matrix*

This section attempts to theoretically define one class of near-optimal binary random matrix in Theorem 4, by comparing the RIPs of LDPC matrices with different sparsities/degrees.

Theorem 4 (Near-optimal binary random matrix): The largest degree d_{max} of LDPC matrix $A(N, M, d_{max})$ with $g > 4$, defines the near-optimal binary random matrix for compressed sensing, which obtains nearly the best RIP (or least RIC-δ_k) with as few nonzero entries as possible.

Proof: The near-optimal LDPC matrix $A(N, M, d_{max})$ is derived by comparing itself with other LDPC matrix $A(N, M, d < d_{max})$ with $g > 4$ and LDPC matrix with $g = 4$.

1) If $d \leq d_{max}$,
   - *compared to LDPC matrix with* $g > 4$:

     with RIP-1 or RIP-2, it is easy to derive that the RIC-δ_k decreases as the degree d increases. Thus, LDPC matrix with $g > 4$ achieves best RIP as $d = d_{max}$.
   - *compared to LDPC matrix with* $g = 4$:

     by comparing RIP-1 and RIP-3, it is easy to derive that $\frac{3k-2}{4d+k-2} < \frac{(3k-2)s}{(k-2)s+4d}$, if $2 \leq s \leq d-1$ and $3 \leq d \leq M/2$ (note that, according to $1 + d(\frac{dN}{M} - 1) \leq N$ [19], one can approximately derive $d_{max} < \sqrt{M} < M/2$). This indicates that the RIC-δ_k of LDPC with $g > 4$ is less than that of LDPC matrix with $g = 4$. Thus, LDPC matrix with $g > 4$ and d_{max} still holds better RIP, if $d \leq d_{max}$.

2) If $d > d_{max}$,
   - *compared to LDPC matrix with* $g = 4$:

     by comparing RIP-1 and RIP-3, LDPC matrix (M, N, d_{max}) with $g = 4$ holds better RIP in the following two cases: *first*, if $d_{max} < d \leq M/2$, let $\frac{3k-2}{4d_{max}+k-2} \leq \frac{(3k-2)s}{(k-2)s+4d}$, derive that $d_{max} \geq d/s$; *second*, if $M/2 < d \leq M - 2$, let $\frac{3k-2}{4d_{max}+k-2} \leq \frac{(3k-2)s+(k-2)(M-2d)}{(k-2)s-(M-2d)k+2M}$, approximately derive that $d_{max} \geq \frac{(k+1)(2d-M)}{6s+2(2d-M)}$.

■

There are three major factors rendering the *near-optimal* rather than *optimal* property of Theorem 4. ***First***, as stated in the *proof*, when $d > d_{max}$, the near-optimal matrix $A(M, N, d_{max})$ is derived under two constraints: $d_{max} \geq d/s$ and $d_{max} \geq \frac{(k+1)(2d-M)}{6s+2(2d-M)}$. For the first constraint, it is reasonable to conjecture that the better LDPC matrix with $g = 4$ and $d > sd_{max}$, dose not exist, since it is contrast to the fact that one can derive at most $d = d_{max}$ when $s = 1$, in terms of the definition d_{max} for LDPC matrix with $g > 4$ in Theorem 4. And for the second constraint, unfortunately it is hard to determine whether the ideal d for better LDPC matrix with $g = 4$ exists or not with the complex formula. Nevertheless, note that, the second constraint is derived from $d > M/2$, which is beyond our practical interest for *sparse* matrix. ***Second***, it is necessary to recall that both RIP-1 and RIP-3 are derived on some special assumptions, which cannot be well satisfied as $|T|$ increases. Specially, as the former states, for LDPC matrix with $g = 4$, the practical RIP may be better than RIP-3, since empirically it is hard for the practical matrix to obey the assumption well as $|T|$ increases. ***Third***, it should be noted that this paper only investigates the *regular* LDPC matrix, which is expected to share same number of nonzero entries between columns or between rows. In fact, Theorem 4 could also be roughly extended to *irregular* LDPC matrix with $g > 4$, by proving that the larger average degree d implies better RIP. In practice, compared to the proposed near-optimal LDPC matrix, irregular LDPC matrix usually can achieve a better RIP thanks to its larger average degree [12].

IV. Simulation results

A. Simulation setup

To verify Theorem 4, this section first studies the performance of LDPC matrix over varying degrees d. Considering the sparsity of input signal is usually uncertain and the noises are also inevitable in real applications, LDPC matrix is also evaluated under above two cases. For comparison, traditional binary random matrix of varying degrees and Gaussian matrix are also tested.

The near-optimal binary random matrix, LDPC matrix $A(200, 400, d_{max} = 7)$ is constructed with PEG algorithm. As a *suboptimal* greedy algorithm, PEG is suitable to explore the largest d of LDPC matrix with $g > 4$. To obtain near-optimal sparsity by comparison, LDPC matrices $A(200, 400,\ 1 \leq d < 7)$ with $g > 4$ and LDPC matrix $A(200, 400,\ 7 < d \leq 100)$ with $g = 4$ are also constructed with PEG algorithm. Moreover, note that this paper cannot provide the examples of LDPC matrix with $g = 4$ and some given $\mu = s/d$, where $2 \leq s \leq d-1$, since there is no construction algorithms as section II states. As particular examples, traditional binary random matrices with $1 \leq d \leq 100$ nonzero entries randomly and uniformly distributed in each column, denoted as $R(M, N, d)$ with the largest correlation $\mu = s/d$ and $s = d - 1$, are proposed for comparison.

Existing solution algorithms to Eq. (2) can be roughly categorized to two classes: convex optimization algorithms (basis pursuit) [28] and greedy algorithms, like orthogonal matching algorithm (OMP) [29] [30]. Since the experiments are developed based on performance comparison rather than pursuing best performance, OMP is used here for much faster simulations. The input sparse signal is generated from $N(0, 1)$ and then normalized. All simulation results are averaged after 10^4 iterations, and both binary random matrix and Gaussian matrix are randomly generated each iteration. The correct recovery rates are measured with $1 - ||\hat{x} - x||_2/||x||_2$.

B. Near-optimal property in both performance and sparsity

TABLE I

THE LARGEST SPARSITY LEVEL k THAT CAN BE RECOVERED WITH PROBABILITY LARGER THAN 99%, FOR LDPC MATRIX $A(200, 400, d)$ AND BINARY RANDOM MATRIX $R(200, 400, d)$.

d		1	2	3	4	5	6	**7**	8	9	10	11	12	13	14	15	20	30	40	50	100
k	LDPC	0	29	70	75	78	80	**81**	**83**	**83**	81	80	79	78	78	77	75	74	48	26	2
	Random	0	0	55	69	73	75	**76**	76	76	76	76	76	76	76	76	76	76	76	76	76

In Table 1, the largest sparsity level k that ensures recovery rates larger than 99% are presented for LDPC matrix binary random matrix with varying d. At the same time, the largest k for Gaussian matrix of size $(200, 400)$ is also derived as 76. Obviously, with same error tolerance, larger k indicates better sensing matrix. As it is expected, the near-optimal binary matrix, LDPC matrix $A(200, 400, d_{max} = 7)$ achieves nearly the best performance with $k = 81$, larger than $k = 76$ for both binary random matrix and Gaussian matrix, while slightly worse than $k = 83$ for LDPC matrices with $d \in \{8, 9\}$. Recall that, since PEG algorithm attempts to reduce the overlap rates of nonzero positions of distinct columns, LDPC matrices with $g = 4$ constructed with PEG algorithm tends to take correlation value $1/d$ with higher probability rather than $2/d$ or others, as d is slightly larger than d_{max}. Therefore, the performance gain of LDPC matrix with $g = 4$ and $d \in \{8, 9\}$ over near-optimal LDPC matrix with $g > 4$ and $d_{max} = 7$, can be explained by the fact that its correlation values take $1/d$ $(< 1/d_{max})$ with much higher probability rather than s/d $(> 1/d_{max})$, $2 \leq s \leq d - 1$. In this sense, this kind of LDPC matrix can be roughly regarded as LDPC matrix with $g > 4$ and $d > d_{max}$, and so the better RIP can be reasonalby conjectured with Theorem 4. Empirically, as Table 1 shows, the performance LDPC matrix constructed with PEG algorithm, begins to degrade as the difference $d - d_{max} > 2$ [12]. By this observation, one can practically construct better LDPC matrix than the near-optimal LDPC matrix with PEG algorithm.

C. *Performance over input signals of varying sparsity or noises*

In Figure 2, near-optimal LDPC matrix is compared with binary random matrix $R(200, 400, 7)$ and Gaussian matrix over varying sparsity levels k. Binary random matrix and Gaussian matrix present comparative performance, while LDPC matrix significantly outperforms them. Note that the binary random matrix with $d = 7$ has achieved its best performance as shown in Table 1. Similar results are also observed in Figure 3, where the normalized sparse signal of $k = 40$ is perturbed by Gaussian noise $N(0, \sigma^2)$.

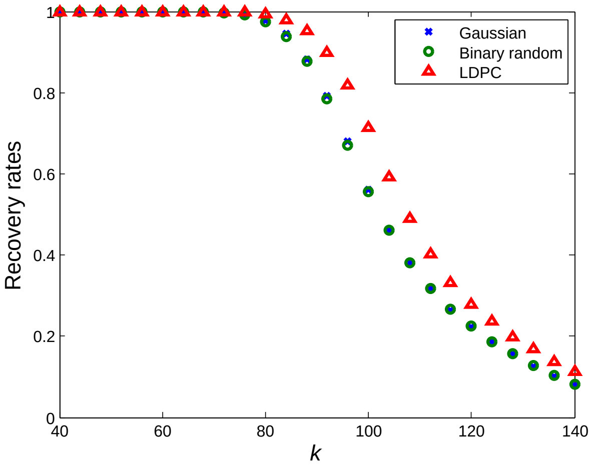


Fig. 2. The recovery rates of input signals with varying sparsity level k, for LDPC matrix $A(200, 400, 7)$, binary random matrix $R(200, 400, 7)$ and Gaussian matrix of size $(200, 400)$.

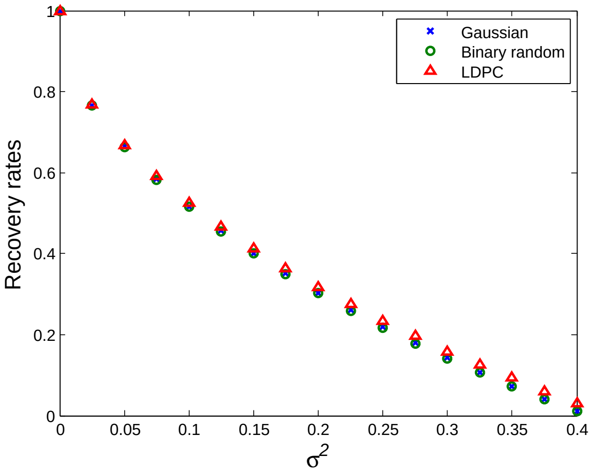


Fig. 3. The recovery rates of normalized input signals with Gaussian noise $N(0, \sigma^2)$, for LDPC matrix $A(200, 400, 7)$, binary random matrix $R(200, 400, 7)$ and Gaussian matrix of size $(200, 400)$.

V. Conclusion

This paper has theoretically defined one class of near-optimal $\{0, 1\}$ binary random matrix with lowest sparsity to achieve nearly the best RIP. And the simulation results also verify that the near-optimal binary matrix constructed with PEG algorithm, indeed exists with better performance over traditional binary random matrix as well as Gaussian matrix. In future, it is attractive to search the largest degree d_{max} for near-optimal LDPC matrix in both theory and practice.

Appendix A
Proof of Theorem 1

Proof referring to [26]: As formula (6) discloses, the solution to RIC-δ_k can be reformulated as the pursuit for the extreme eigenvalues of symmetric matrix $A'_T A_T \subset \{0, 1, 1/d\}^{k \times k}$, $|T| = k$. Customarily, as the former states, we denote by $\lambda_1(A'_T A_T) \geq \ldots \geq \lambda_k(A'_T A_T)$ the eigenvalues of $A'_T A_T$.

1) Let $B = A'_T A_T - I \subset \{0, 1/d\}^{k \times k}$, then $B_{ii} = 0$ and $B_{ij, i \neq j} = 0$ or $1/d$.

   Let normalized $x = (x_1, \ldots, x_k)'$ be the eigenvector corresponding to $\lambda_k(B)$. Then the minimal eigenvalue can be formulated as

$$\lambda_k(B) = x'Bx = \mathbb{1}'[B \circ (xx')]\mathbb{1}$$

   where $\circ$ denotes the Hadamard product and $\mathbb{1} = (1, \ldots, 1)' \in \mathbb{R}^k$. Since B is symmetric, by simultaneous permutations of the rows and columns of B, we can suppose $x_i \geq 0$ for $i = 1, \ldots, n$ and $x_i < 0$ for $i = n+1, \ldots, k$, and then xx' is divided into four parts:

$$xx' = \begin{bmatrix} X_{n \times n} & X_{n \times (k-n)} \\ X_{(k-n) \times n} & X_{(k-n) \times (k-n)} \end{bmatrix}$$

   where the entries in $X_{n \times n}$ and $X_{(k-n) \times (k-n)}$ are nonnegative, while the entries in $X_{n \times (k-n)}$ and $X_{(k-n) \times n}$ are nonpositive. Further, define a novel matrix $\tilde{B}$ of same size with B

$$\tilde{B} = \begin{bmatrix} 0 \times \mathbb{1}_{n \times n} & \frac{1}{d} \times \mathbb{1}_{n \times (k-n)} \\ \frac{1}{d} \times \mathbb{1}_{(k-n) \times n} & 0 \times \mathbb{1}_{(k-n) \times (k-n)} \end{bmatrix}$$

   where $\mathbb{1}_{a \times b}$ is an $a \times b$ matrix with all entries equal to 1. It is easy to deduce that

$$\lambda_k(\tilde{B}) = \min\{y'\tilde{B}y : \|y\| = 1\} \leq x'\tilde{B}x \leq x'Bx = \lambda_k(B).$$

   Since the rank of $\tilde{B}$ is at most 2, it has at most two nonzero eigenvalues. Considering the trace and the Frobenius norm, we have

$$\lambda_k(\tilde{B}) = -\sqrt{\frac{n(k-n)}{d^2}},\ 0 \leq n \leq k.$$

   If k is even, $\lambda_k(\tilde{B}) \geq -\frac{k}{2d}$, with '=' at $n = k/2$.

   If k is odd, $\lambda_k(\tilde{B}) \geq -\frac{\sqrt{k^2-1}}{2d}$, with '=' at $n = (k-1)/2$ or $n = (k+1)/2$.

   Then $\lambda_k(B) \geq \lambda_k(\tilde{B}) \geq -\frac{k}{2d}$, with the limitation attained at k is even and $n = k/2$.

   So, we have the minimum eigenvalue $\lambda_k(A'_T A_T) \geq 1 - \frac{k}{2d}$.

2) Let $C = A'_T A_T - \frac{d-1}{d} \times I$, then $C_{ii} = 1/d$ and $C_{ij, i \neq j} = 0$ or $1/d$.

Let normalized $x = (x_1, \ldots, x_k)'$ be the eigenvector corresponding to $\lambda_1(C)$. By simultaneous permutations of C and x, we can suppose $x_i \geq 0$ for $i = 1, \ldots, n$ and $x_i < 0$ for $i = n+1, \ldots, k$, and the maximal eigenvalue is formulated as

$$\lambda_1(C) = x'Cx = \mathbb{1}'[C \circ (xx')]\mathbb{1}.$$

Further define

$$\tilde{C} = \begin{bmatrix} \frac{1}{d} \times \mathbb{1}_{n\times n} & 0 \times \mathbb{1}_{n\times(k-n)} \\ 0 \times \mathbb{1}_{(k-n)\times n} & \frac{1}{d} \times \mathbb{1}_{(k-n)\times(k-n)}, \end{bmatrix}$$

then

$$\begin{aligned} \lambda_1(\tilde{C}) &= \max\{y'\tilde{C}y : \|y\| = 1\} \geq x'\tilde{C}x \geq x'Cx \\ &= \lambda_1(C). \end{aligned}$$

Since the rank of $\tilde{C}$ is at most 2, it has at most two nonzero eigenvalues. Considering the trace and the Frobenius norm, we have

$$\lambda_1(\tilde{C}) = \frac{k + |k - 2n|}{2d}.$$

Then $\lambda_1(C) \leq \lambda_1(\tilde{C}) \leq \frac{k}{d}$, with '=' at $n = 0$ or $n = k$. And further derive

$$\lambda_1(A_T'A_T) = \lambda_1(C) + \frac{d-1}{d} \leq \frac{k+d-1}{d}$$

3) Finally, we can deduce

$$\delta_k = \frac{\lambda_1(A_T'A_T) - \lambda_k(A_T'A_T)}{\lambda_1(A_T'A_T) + \lambda_k(A_T'A_T)} = \frac{3k-2}{4d+k-2},$$

with $\frac{\lambda_1(A_T'A_T)}{\lambda_k(A_T'A_T)} = \frac{1+\delta_k}{1-\delta_k}$ [31].

■

Appendix B
Proof of Theorem 2

Proof: To derive the extreme eigenvalues of $A_T'A_T$, we first search the extreme eigenvalues of

$$B = (A_T'A_T - I)$$

where I is an identity matrix. And clearly B is a symmetric matrix of the diagonal elements equal to 0, and the off-diagonal elements equal to 1 with property ρ and 0 with property $1 - \rho$.

Further, suppose [32]

$$Q = \frac{1}{\sqrt{\rho(1-\rho)}}(B - \rho\mathbb{1})$$

where $\mathbb{1}$ is a all-ones matrix. Then Q has entries with mean zero and variance one. With *semicircle law* [27] , the extreme eigenvalues $\frac{1}{\sqrt{k}}Q$, $k = |T|$, can be approximated as

$$-2 \leq \lambda(\frac{1}{\sqrt{k}}Q) \leq 2$$

namely,

$$-2\sqrt{k\rho(1-\rho)} \leq \lambda(B - \rho\mathbb{1}) \leq 2\sqrt{k\rho(1-\rho)},$$

if $k \to \infty$ [33].

With *cauchy interlacing inequality* [34], we can further derive that

$$\lambda_i(B - \rho\mathbb{1}) \leq \lambda_i(B) \leq \lambda_{i-1}(B - \rho\mathbb{1})$$

for $1 < i \leq k$, if $B - \rho\mathbb{1}$ and $\rho\mathbb{1}$ are Hermitian matrices, and $\rho\mathbb{1}$ is positive semi-definite and has rank equal to 1. As a result, it is easy to derive that

$$\lambda_2(B) \leq \lambda_1(B - \rho\mathbb{1}) \leq 2\sqrt{k\rho(1-\rho)}$$

and

$$\lambda_k(B) \geq \lambda_k(B - \rho\mathbb{1}) \geq -2\sqrt{k\rho(1-\rho)}$$

As for $\lambda_1(B)$ [2], it is known that [36]

$$\lambda_1(B) \approx k\rho + 1$$

In this sense, the extreme eigenvalues of $A_T' A_T$ can be approximately formulated as

$$\lambda_1(A_T' A_T) = \lambda_1 B + 1 \leq k\rho + 2$$

and

$$\lambda_k(A_T' A_T) = \lambda_k B + 1 \geq -2\sqrt{k\rho(1-\rho)} + 1$$

Therefore, the RIC of $A_T' A_T$ can also be deduced

$$\delta_k = \frac{\lambda_1 - \lambda_k}{\lambda_1 + \lambda_k} = \frac{k\rho + 2\sqrt{k\rho(1-\rho)} + 1}{k\rho - 2\sqrt{k\rho(1-\rho)} + 3}$$

■

[2] In [35], it is proved that $\lambda_1(B) \approx k\rho$, as $k\rho$ is sufficiently large.

Appendix C
Proof of Theorem 3

The proof is similar to that for Theorem 1 in Appendix A. So in the following we just give a sketch.

Proof:

1) If $3 \leq d \leq M/2$, $[A'_T A_T]_{ii} = 1$ and $[A'_T A_T]_{ij,i\neq j} \in \{0, \ldots, s/d\}$, $2 \leq s \leq d-1$, for $i, j = 1, \ldots, k$.
   a) Let $B = A'_T A_T - I$,
   $$\lambda_k(B) \geq \begin{cases} -sk/2d & if\ k\ is\ even \\ -s\sqrt{k^2-1}/2d & if\ k\ is\ odd \end{cases}$$
   then $\lambda_k(A'_T A_T) = 1 + \lambda_k(B) \geq 1 - \frac{sk}{2d}$.
   b) Let $C = A'_T A_T - (1 - \frac{s}{d})I$, derive $\lambda_1(C) \leq ks/d$, then $\lambda_1(A'_T A_T) \leq \frac{(k-1)s+d}{d}$.
2) if $M/2 < d \leq M - 1$, $[A'_T A_T]_{ii} = 1$ and $[A'_T A_T]_{ij,i\neq j} \in \{(2d - M)/d, \ldots, s/d\}$, $2d - M \leq s \leq d - 1$, for $i, j = 1, \ldots, k$.
   a) Let $B = A'_T A_T - (1 - \frac{2d-M}{d})I$, derive $\lambda_k(B) \geq \begin{cases} \frac{k(2d-M-s)}{2d} & if\ k\ is\ even \\ \frac{k(2d-M)-\sqrt{(2d-M)^2-(k^2-1)s^2}}{2d} & if\ k\ is\ odd \end{cases}$
   further derive $\lambda_k(B) \geq -\frac{k(2d-M-s)}{2d}$, and then we have that $\lambda_k(A'_T A_T) \geq \frac{k(2d-M-s)+2(M-d)}{2d}$
   b) Let $C = A'_T A_T - (1 - \frac{s}{d})I$, derive $\lambda_1(C) \leq ks/d$, and then it follows that $\lambda_1(A'_T A_T) \leq \frac{(k-1)s+d}{d}$.
3) Finally, with $\delta_k = \frac{\lambda_1 - \lambda_k}{\lambda_1 + \lambda_k}$, derive

$$\delta_k = \begin{cases} \frac{(3k-2)s}{(k-2)s+4d} & if\ 3 \leq d \leq \frac{M}{2} \text{ and } 2 \leq s \leq d-1 \\ \frac{(3k-2)s+(k-2)(M-2d)}{(k-2)s-(M-2d)k+2M} & if\ \frac{M}{2} < d \leq M-2 \text{ and } 2d - M \leq s \leq d-1 \end{cases}$$

■

References

[1] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.

[2] S. Foucart, "A note on guaranteed sparse recovery via l_1-minimization," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 97 – 103, 2010.

[3] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406 –5425, Dec. 2006.

[4] D. Achlioptas, "Database-friendly random projections: Johnson–Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, 2003.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[6] M. Akcakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2443 –2450, june 2008.

[7] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *Journal of Complexity*, vol. 23, no. 4-6, pp. 918 – 925, 2007.

[8] S. Howard, A. Calderbank, and S. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order reed-muller codes," in *42nd Annual Conference on Information Sciences and Systems (CISS 2008)*, march 2008, pp. 11 –15.

[9] L. Applebaum, S. D. Howard, S. Searle, and R. Calderbank, "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery," *Applied and Computational Harmonic Analysis*, vol. 26, no. 2, pp. 283 – 290, 2009.

[10] D. Li, X. Liu, S. Xia, and Y. Jiang, "A class of deterministic construction of binary compressed sensing matrices," *Journal of Electronics (China)*, vol. 29, pp. 493–500, 2012.

[11] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21 –28, Jan 1962.

[12] W. Lu, K. Kpalma, and J. Ronsin, "Sparse binary matrices of LDPC codes for compressed sensing," in *Data Compression Conference (DCC), 2012*, april 2012, p. 405.

[13] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269 –280, jan. 2010.

[14] M. Akcakaya, J. Park, and V. Tarokh, "Low density frames for compressive sensing," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, march 2010, pp. 3642 –3645.

[15] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," in *IEEE Information Theory Workshop, ITW '07*, sept. 2007, pp. 414 –419.

[16] S. Sarvotham, D. Baron, and R. Baraniuk, "Sudocodes: Fast measurement and reconstruction of sparse signals," in *IEEE International Symposium on Information Theory*, july 2006, pp. 2804 –2808.

[17] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," *MIT-CSAIL Technical Report*, 2008.

[18] A. Dimakis, R. Smarandache, and P. Vontobel, "LDPC codes for compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3093 –3114, may 2012.

[19] X.-Y. Hu, E. Eleftheriou, and D. Arnold, "Regular and irregular progressive edge-growth Tanner graphs," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 386 –398, jan. 2005.

[20] D. Needell and J. A. Tropp, "Cosamp: iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec. 2010.

[21] A. M. Tillmann and M. E. Pfetsch, "The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing," *ArXiv e-prints*, May 2012.

[22] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *ArXiv e-prints*, Apr. 2012.

[23] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" *SIAM Rev.*, vol. 53, no. 1, pp. 105–125, Feb. 2011.

[24] M. Rudelson and R. Vershynin, "Non-asymptotic theory of random matrices: extreme singular values," in *International Congress of Mathematicans*, 2010.

[25] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2011.

[26] X. Zhan, “Extremal eigenvalues of real symmetric matrices with entries in an interval,” *SIAM Journal on Matrix Analysis and Applications*, vol. 27, no. 3, pp. 851–860, 2005.

[27] L. Pastur, “On the spectrum of random matrices,” *Theoretical and Mathematical Physics*, vol. 10, pp. 67–74, 1972.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambrige university press, March 2004.

[29] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, nov 1993, pp. 40 –44 vol.1.

[30] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transaction on Information Theory*, vol. 53, pp. 4655–4666, 2007.

[31] S. Foucart and M.-J. Lai, “Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 \leq q < 1$,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395 – 407, 2009.

[32] L. V. Tran, V. H. Vu, and K. Wang, “Sparse random graphs: Eigenvalues and eigenvectors,” *Random Structures & Algorithms*, vol. 42, no. 1, pp. 110–134, 2013.

[33] Z. Füredi and J. Komlós, “The eigenvalues of random symmetric matrices,” *Combinatorica*, vol. 1, pp. 233–241, 1981.

[34] T. Tao and V. Vu, “Random matrices: Universality of local eigenvalue statistics,” *Acta Mathematica*, vol. 206, pp. 127–204, 2011.

[35] T. Ando, Y. Kabashima, H. Takahashi, O. Watanabe, and M. Yamamoto, “Spectral analysis of random sparse matrices,” *IEICE Transactions*, pp. 1247–1256, 2011.

[36] Y. Kabashima, H. Takahashi, and O. Watanabe, “Cavity approach to the first eigenvalue problem in a family of symmetric random sparse matrices,” *Journal of Physics: Conference Series*, vol. 233, no. 1, p. 012001.